Online Infrastructure Dependency Detection and Tracking

István Szombath

IT infrastructures providing vital business services are becoming more and more distributed and heterogeneous. System management has to assure the appropriate quality of services and keep resource usage at a reasonable level. Structural information, especially the dependencies between IT components, is vital to system management. Without the dependency information it is not feasible to determine the impact of IT component faults on business services. Dependency information has foremost importance in adaptive architectures, like dynamic reconfiguration based self healing systems. For instance configuration consolidation (i.e., reallocation of servers) in a virtualized infrastructure (cloud) is only feasible when the dependency information is known. Due to the widespread use of adaptive architectures, gathering information on the structure of the system becomes increasingly important for practical system management.

Current state of the art shows that with passive observation of network communication (e.g., with NetFlow) reconstruction of the IT infrastructure model is feasible. The reconstructed model represents servers and the communication of servers. Important dependencies between servers (so called service dependency, e.g., a dependency between a web server and a database) can be identified from the reconstructed model, for example using the method presented in [1].

The model (e.g., a labeled graph) of the infrastructure can be very complex and it could change rapidly. However even a huge enterprise class IT infrastructure can be described with only a few types of high level service patterns, such as 3 tier architectures, backups, authentication and mailing solutions, etc. Thus the most part of the infrastructure graph are covered by these service patterns [2]. The drawback and challenges of this approach are the computational complexity of pattern matching, and the typical patterns needs to be collected manually.

Our method builds a labeled graph from passive observation of network communication that represents the IT infrastructure and updates it online. A method is also worked out to identify and track the existence of typical patterns of the IT infrastructure, e.g. a 3 tier architecture. Typical service patterns can be set, and the pattern matcher [3] identifies the matches in the model online. This means upon model update new matches may be found or already found matches can became obsolete. A proof of concept is also presented to collect typical service patterns automatically using graph clustering. The engine is capable to evaluate the patterns in real time, even in large scale. The approach is verified using communication logs of a real IT infrastructure. The framework is also capable to propagate the discovered dependency information to an enterprise class system management model repository (IBM Tivoli CCMDB).

References

[1] Andreas Kind, Dieter Gantenbein, Hiroaki Etoh, *Relationship Discovery with NetFlow to Enable Business-Driven IT Management*, 1st IEEE/IFIP Int. Workshop on Business- riven IT Management, 2006.

[2] Thomas Karagiannis,Konstantina Papagiannaki, Michalis Faloutsos, *BLINC: Multilevel Traffic Classification in the Dark*, ACM SIGCOMM Computer Communication Review, vol. 35, no. 4, pp. 229-240, October, 2005.

[3] Ráth, I., Bergmann, G., Ökrös, A., and Varró, D. *Live Model Transformations Driven by Incremental Pattern Matching*. In Proceedings of the 1st international Conference on theory and Practice of Model Transformations (Zurich, Switzerland, July 01 - 02, 2008).